DOCUMENT RESUME

ED 232 590 HE 016 514

AUTHOR Greene, James E., Jr.; And Others
TITLE Using Administrative Data as Unobtrusive Indicators of Teaching Performance. AIR 1983 Annual Forum Paper.
PUB DATE May 83
NOTE 18p.; Paper presented at the Annual Forum of the Association for Institutional Research (23rd, Toronto, Ontario, May 23-26, 1983).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Courses; *Evaluation Methods; *Faculty Evaluation; Higher Education; Institutional Research; Reliability; *Student Evaluation of Teacher Performance; *Teacher Effectiveness
IDENTIFIERS *AIR Forum; *Course Selection

ABSTRACT

In order to address the issue of the reliability and possible bias of students' evaluation of the performance of college teachers, an unobtrusive indicator of teacher performance was investigated. The indicator was the degree to which teachers develop a following among students, measured by how often students return to teachers for additional courses. The model uses 5 years of student course and grade data at a large, state university, which entailed the records of 50,000 students, the one-half million courses they took, and 2,275 teachers. The findings suggest that the measure of students taking additional courses from faculty members can be a valid indicator of teaching effectiveness, and can be used along with students' ratings. The ability to use this type of administrative data to augment existing systems of evaluating teacher effectiveness is advocated, since diversity of measurement not only reinforces reliability, but can enhance the effect of valid and unbiased teacher performance indicators. It is concluded that such data, when used in conjunction with other evaluative methods, can be a valuable resource. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

#38

ED232590

USING ADMINISTRATIVE DATA AS UNOBTRUSIVE INDICATORS OF TEACHING PERFORMANCE

James E. Greene, Jr., Registrar
Registrar's Office
Georgia State University
University Plaza
Atlanta, Georgia 30303
(404) 658-2386

James E. Prather, Senior Research Associate
Office of Institutional Planning
Georgia State University
University Plaza
Atlanta, Georgia 30303
(404) 658-2570

Joseph E. Sturgeon, Management Information Specialist
Registrar's Office
Georgia State University
University Plaza
Atlanta, Georgia 30303
(404) 658-2389

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

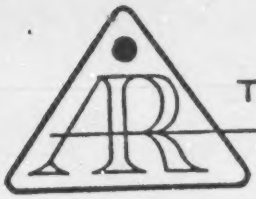
AIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE016514

Contributed Paper
Association for Institutional Research Forum
May 23-26, 1983
Toronto, Ontario, Canada

Printed in U. S. A.



THE ASSOCIATION FOR INSTITUTIONAL RESEARCH

This paper was presented at the Twenty-Third Annual Forum of the Association for Institutional Research held at the Sheraton Centre in Toronto, Ontario, May 23-26, 1983. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum papers.

D. R. Coleman, Chairman
Forum Publication
Advisory Committee

Using Administrative Data as Unobtrusive Indicators of Teacher Performance

Few issues in colleges and universities raise greater debate or anxieties than the question of students evaluating classroom teaching. Research in this area has been appearing in academic journals with increasing frequency for over a decade. Many faculty members and academic administrators have found themselves choosing ideological sides in this issue. The debate includes both policy questions and measurement problems. The policy questions such as interference with academic freedom and the ability of students to judge quality teaching do not hinge on measurement problems. However, numerous other policy issues are dependent on arriving at some consensus solution on various measurement and methodological problems. These measurement problems include questions of reliability, validity, and bias in the presence of students' judging teaching effectiveness. The great majority of research on students' evaluation of teaching has revolved around the presence of bias in students' rating of teachers and how to achieve a minimum of bias.

Research on student evaluation has introduced various measures of teaching effectiveness. While self evaluation and peer evaluation have been used (with mixed results), the primary source of measurements to determine teacher effectiveness has been anonymous ratings by students. Predictably, this system of students' secret scoring of their instructors has itself generated great controversy, especially in the areas of reliability, validity, and bias. The question of reliability of student ratings has been addressed with a certain amount of success. The validity of student ratings of teacher performance and the ratings' freedom from bias are not yet generally accepted.

Moreover, unobtrusive measurement of teacher effectiveness is itself a difficult task. There are well-founded apprehensions that the measurement process so affects students and teachers that any reading of teacher effectiveness is necessarily distorted. Research literature has recorded that factors outside the direct influence of the instructor, such as whether the course taught is required or is an elective, and how large the class enrollment is, have an effect on student ratings of teacher performance.

The great need of administrators to accurately gauge teacher performance, combined with problems of reliability, validity, and bias in current measures, enhances the need to find alternate measures of teaching performance. In this paper, a proposed indicator of teacher performance is introduced. This indicator is based on observable student behavior, rather than on the difficult to determine factors that constitute "opinion," whether student, peer, or administrator. The observable student behavior that is proposed as an indicator is the number of times students return to a teacher for additional courses. In large, diverse colleges and universities, students are given a great deal of freedom in choosing their programs of study and, in many cases, choosing a particular schedule. This means that student behavior can exhibit a pattern of avoiding or seeking certain teachers. In addition, those special teachers that stimulate interest in their disciplines and encourage intellectual development should be indicated by how often they are sought out by the students who are their followers or proteges. Furthermore, this indicator makes use of existing administrative data and has the advantage of being totally unobtrusive.

Review of Literature

Many studies have been conducted to determine what factors influence or bias students' ratings of instructors. The effect of instructor-student congruence of attitudes on instructor ratings was investigated by Hofman and Kremer (1980). Multiple regression analysis showed that almost 50% of the variance in evaluations could be accounted for by attitudes. Instructors' attitudes as perceived by students were the best predictors of evaluations, and a positive course evaluation was associated with students attributing to instructors attitudes preferred by the students themselves.

There is evidence that the personality characteristics of the instructor have an effect on instructor ratings. In an analysis of student written answers to open-ended evaluation questions, Braskamp, Ory, and Pieper (1981) found that only one fourth of the comments pertained to the instructor's pedogogical skills, while approximately half of the comments were about the instructor himself. Abrami, Perry and Leventhal (1982) examined the relationship between student personality characteristics, teacher ratings, and student achievement. While they found no consistent relationship between ratings and student personality characteristics, the perceived personality characteristics of instructors were related to ratings of teacher effectiveness. They also found that teacher effects on ratings appeared significantly greater than teacher effects on achievement.

Other studies have looked at student and course variables to explain instructor ratings. Marsh and Overall (1980) emphasize the importance of considering multiple criteria in the evaluation of teaching effectiveness. They found that both cognitive and affective criteria of effective teaching were correlated with different components of student ratings but were uncorrelated with each other. Marsh (1980) examined the relationship between student

evaluations and a set of 16 background characteristics. Multivariate analysis suggested that up to 14% of the variance in student ratings could be accounted for by the 16 variables. Favorable ratings were correlated with prior subject interest, higher expected grades, higher levels of workload difficulty, and a higher percentage of students taking the course for general interest only.

Overall and Marsh (1980) investigated the relative contribution of course level (undergraduate versus graduate), course type (accounting, economics, etc.), and the specific instructor on students' evaluation of instruction. They found that the variance which could be attributed to the specific instructor was much greater than that due to course level or course type. Their findings indicate that who teaches a course is relatively more important than the particular course or the level at which it is taught.

Much has been written on the relationship between grades and instructor ratings. The results of a study by DuCette and Kenny (1982) indicate that the ratings a student gives to an instructor and a course are somewhat determined by the grade the student expects to obtain in the course. Abrami, Dickens, Perry, and Leventhal (1980) found that teacher standards for assigning grades had significant effects on student ratings, but concluded that, when using ratings to make gross distinctions between teachers, the size of the differences was relatively unimportant. Howard and Maxwell (1980) argue against the grading leniency bias model (the idea that instructors who are easy graders receive high ratings simply because they are easy graders). They describe two studies suggesting that other variables, especially student motivation, may be more important in determining instructor evaluations. Howard and Maxwell (1982) replicated their earlier study and found no evidence that a grades-influencing-satisfaction interpretation was more likely than its opposite.

There has been a controversy in the literature concerning the validity and use of instructor ratings for faculty personnel decisions. The findings of Marsh (1982) demonstrated agreement between students and instructors on evaluations of teaching effectiveness. According to Marsh, his findings support the validity of student ratings. Dowell and Neal (1982) provide a review of studies which have attempted to link student ratings to student learning as a way of validating student ratings as a measure of teaching ability. Their evidence suggests that the validity of student ratings is quite variable, and is at best only modest. They recommend that student ratings be used with great caution in the processes of faculty review and decision making. Hills (1974) analyzed articles and reviews of past literature on ratings to determine how student ratings were being used, their relationship to teaching effectiveness, and the appropriateness of their use in making personnel decisions. He concluded that student ratings of faculty could not be trusted when determining pay, promotion, and tenure.

Methodology

Data

The data necessary to determine a given teacher's rate of student repeats were obtained from the student data base. A total of 50,000 students who have taken half a million courses was matched with 2275 teachers for the period of winter quarter 1978 through winter quarter 1982. The average number of repeats per instructor was .34, with a standard deviation of .56. The modal rate of student repeats was zero and the distribution was quite skewed toward zero, but with a very long tail that extended to a rate of repeating of over four.

Methods

The analysis was done by aggregating various academic and demographic characteristics of students for each teaching faculty member and obtaining means for the number of courses taught. The variables included in the analysis measured such characteristics of instructional mode as the percent of lectures, seminars, theses, and all "other" courses being taught by an instructor. The average class size for each instructor was included. Instructional level was measured as the percent of graduate, upper, or lower division courses taught. The student demographic characteristics included the percent of minority males, minority females, nonminority males, and nonminority females. The measures of student academic performance were the average GPA, average grade, percent of students withdrawing from courses, average rate of attendance for the five-year period, and number courses taken previously. A number of variables were included to control for both experience and teaching load of the instructors, such as number of courses and quarters taught and the inverse of the number of courses and quarters taught.

Findings

Multiple regression was the technique chosen to explore the influence of various factors upon students' tendencies to repeat a given instructor. The multiple regression for the 2275 teachers is reported in Table 1. The strongest positive influence on the rate of students repeating a given instructor is the average rate of attendance by the students for the five-year period. Other positive factors influencing the rate of student repeats show that as the average GPA of the students taught and the average grade assigned increases, the student repeat rate also increases. Negative factors include percent lectures taught and average class size. This means that the teaching faculty members who teach more lectures and have higher average class sizes experience fewer student repeaters. In addition, as the average of the number of previous courses taken by their students increases, their rate of student repeats decreases. The adjusted R^2 for this regression was .48, indicating that almost half of the variance in the average rate of student repeats can be accounted for by known factors.

For a selected group of instructors in a liberal arts discipline that taught both service and degree courses at the undergraduate and graduate level, data were obtained from students who evaluated the teachers on seven measures of teaching effectiveness. This information was obtained for 52 teaching faculty members. These seven measures of teaching effectiveness include if the instructor: 1) was well-prepared; 2) stimulated student thinking; 3) was actively helpful to students; 4) explained course

Table 1

A Multiple Regression of the Rate of Students Repeating a Teacher with Instructional, Academic, and Demographic Characteristics for all Teachers, 1978-82

Variables	Unstandardized Regression Coefficients	Standardized Regression Coefficients
Average Department Repeats	.07	.28
Instructional Mode:		
%Lectures	-.003	-.16
%Seminars	-.003	-.06
%Thesis	.01	.10
Average Class Size	-.01	-.16
Instructional Level:		
%Graduate	-.0007	-.05
%Upper	-.001	-.07
Student Demographics:		
%Minority Female	.002	.03
%Minority Male	.006	.08
%Non-Minority Female	.0003	.01
Student Academic Performance:		
Average GPA	.0009	.10
Average Grade	.0006	.08
%Withdrawing	-.002	-.04
Attendance Rate	.03	.29
Number of Previous Courses Taken	-.03	-.21
Controls:		
Number Courses Taught	.006	.23
Number Quarters Taught	.007	.08
Inverse of Number of Courses Taken	-.35	-.09
Constant	-.18	
R^2		.49
Adjusted R^2		.48
N		2275

objectives; 5) was fair and impartial in grading; 6) explained difficult material; and 7) if the students felt they learned a great deal. Regressions of the seven items and the total average of these seven items are included in Table 2. Another regression included is that of the average student repeats.

The interpretation of Table 2 emphasizes the extent to which the explanatory variables have differing impacts or weights across the seven items measuring teaching effectiveness. Of particular note is the infuence of student repeats upon the rating items. The strongest weight of student repeats is the "explains difficult material" item. This may well mean that the handling of difficult material is more salient to students in general than other measures of teaching effectiveness. As the percent of lectures taught by faculty members increases, their student ratings consistently decrease. Average class size is negatively associated with all ratings except the "explains difficult material" item. The higher proportion of lower division courses, the higher the ratings, but the effect is negative for graduate and upper divisions. Teaching higher percentages of minority males and females is indicative of lower ratings. The higher the GPA's of the students being taught, the more withdrawals, and the greater the rate of student attendance, the lower the ratings. The more courses previously taken, the more likely it is that the faculty received higher ratings.

The regression of the rate of students repeating the same teacher is also shown in Table 2. The seven items are also entered into this regression. They show the two strongest positive indicators of repeats to be "explains difficult material" and "well-prepared." Strangely, the two

Table 2

Multiple Regressions of Students' Ratings of Teachers and Rate of Students Repeating A Teacher with Instructional, Academic, and Demographic Characteristics for Selected Teachers, 1978-1982

VARIABLES	Well-Prepared	Stimulated Student Thinking	Actively Helpful to Students	Course Objectives Explained	Grades Fair/ Impartial	Explains Difficult Material	Students Learn Great Deal	Average of 7 Items	Average Repeats
Average Student Repeats	.78	.49	.66	.71	.25	1.20	.43	.65	
Instructional Mode:									
% Lectures	.007			.007	-.01		.02	.01	
% Seminars	-.24	-.18		-.09		-.25	-.09	-.11	.11
% Thesis		-.02			-.05				
% Other	.02	.01	.01	.04		.01	.04	.03	-.005
Average Class Size	-.006	-.007		-.009	-.01	.005	-.006	-.005	-.006
Instructional Level:									
% Graduate	-.01	-.01	-.004	-.007	-.008	-.01	-.01	-.01	.002
% Upper	-.003	-.002			-.007		-.003	-.005	.007
% Lower			.006	.008		.009			
Student Demographics:									
% Minority Female	-.03	-.03	-.01	-.02	-.01	-.03	-.02	-.02	-.007
% Minority Male	-.04	-.05	-.06	-.04	-.03	-.08	-.05	-.05	-.005
% Non-minority Female				-.01	-.005				
% Non-minority Male	-.02	-.02	-.001			-.02	-.008	-.01	-.0001
Student Academic Performance:									
Average GPA	-.08	-.60	-.52	-.13	-.25	-.43	-.40	-.34	-.26
Average Grade	.39	.60	.63	.56	.48	.70	.59	.56	.14
% Withdrawing	-.001	-.006	-.008	-.001	-.001	-.01	-.003	-.005	.003
Attendance Rate	-.08	-.07	-.08	-.06	-.04	-.13	-.07	-.08	
No.of Previous Courses Taken	.04		.02	.02	.02	.04	.005	.02	-.01
Controls:									
No. Courses Taught	-.008	-.004	-.008	-.01	-.003	-.01	-.007	-.008	.007
No. Quarters Taught	.01	.009	.02	.03	.01	.02	.01	.02	
Inverse No. Courses Evaluated	.05	-.17	-.26		-.15	-.16	-.12	-.11	
Inverse No. Courses Taught									-1.41
Evaluation Items of Instructor:									
1) Well-prepared									.21
2) Stimulated Students' Thinking									-.27
3) Actively Helpful to Students									-.02
4) Course Objectives Explained									.04
5) Grades Fair/Impartial									-.09
6) Explains Difficult Material									.22
7) Students Learn Great Deal									-.18
Constant	5.34	6.98	5.58	3.71	6.63	6.11	4.16	4.77	.83
R^2	.42	.47	.53	.52	.60	.45	.43	.48	.92
R^2 (Adjusted)	.10	.21	.33	.28	.40	.18	.11	.20	.86
N	52	52	52	52	52	52	52	52	52

strong negative coefficients are "stimulated student thinking" and "student learned a great deal." These negative variables may well be the "suppressor" effects, and this is unclear at this level of analysis.

Of key importance is the question of the validity of the indicators of the relationship between students repeating a given faculty member for another course and the measures of students' evaluations of teaching effectiveness. These validity coefficients (bivariate correlations) are given in Table 3. The strongest correlations are those of approximately .22, which are evidence of a moderate level of validity. Even though the lowest correlation is with the "explains difficult material" item, this item had the strongest regression coefficients in the multiple regressions.

Conclusion and Implications

There has been much debate about the reliability, and bias when students evaluate the performance of teachers in college classrooms. This paper presents an unobtrusive indicator of teacher performance that may prove helpful in this debate. The indicator is the degree to which teachers develop a following among students as measured by how often students return to teachers for additional courses. The indicator can be derived from available administrative data.

The model uses five years of student course and grade data at a large, state university. Using the records of 50,000 students and the one-half million courses they took, the research determines which teachers were repeated for each student. Non-teaching factors are controlled for and student evaluations of teachers from other sources are then correlated with the results.

Table 3

Correlations of Student Rate of Repeating the Same Teacher with Measures of Teaching Effectiveness by Students for Selected Teaching Faculty Members, 1978-1982
N=52

Measures of Teaching Effectiveness	Rate of Students Repeating Same Teacher
1) Well-prepared	.088
2) Stimulated student thinking	.225
3) Helpful to students	.190
4) Explained course objectives	.221
5) Grades fair/impartial	.221
6) Explains difficult material	.028
7) Students learn great deal	.162
Total	.163

The available evidence suggests that the measure of students' repeating faculty members can be a valid indicator of teaching effectiveness. It appears that the repeat measure could well serve as an indicator of teaching effectiveness along with students' ratings. Student ratings and the repeat measure are subject to the same biases. Teacher personality and student motivation and orientation could affect the repeat measure in a manner similar to their effect on student ratings of teacher effectiveness.

The ability to use specific administrative data to augment existing systems of evaluating teacher effectiveness is certainly desirable. Diversity of measurement not only reinforces reliability, it also can enhance the effect of valid and unbiased teacher performance indicators. Administrative data have the potential to both add much-needed information regarding the evaluation of teacher performance and define additional criteria vital to the assessment of teacher effectiveness. It is concluded that such data, when used in conjunction with other evaluative methods, can be a valuable resource.

References

Abrami, P.C., Dickens, W.J., Perry, R.P, & Leventhal, L. Do teacher standards for assigning grades affect student evaluations of instruction? Journal of Educational Psychology, 1980, 72, (1), 107-118.

Abrami, P.C., Perry, R.P., & Leventhal, L. The relationship between student personality characteristics, teacher ratings, and student achievement. Journal of Educational Psychology, 1982, 74, (1), 111-125.

Braskamp, L.A., Ory, J.C., & Pieper, D.M. Student written comments: Dimensions of instructional quality. Journal of Educational Psychology, 1981, 73, (1), 65-70.

Dowell, D.A., & Neal, J.A. A selective review of the validity of student ratings of teaching. Journal of Higher Education, 1982, 53, (1), 51-62.

DuCette, J., & Kenney, J. Do grading standards affect student evaluations of teaching? Some new evidence on an old question. Journal of Educational Psychology, 1982, 72, (3), 308-314.

Hills, J.R. On the use of student ratings of faculty in determination of pay, promotion and tenure. Research in Higher Education, 1974, 2, (4), 317-324.

Hofman, J.E., & Kremer, L. Attitudes toward higher education and course evaluation. Journal of Educational Psychology, 1980, 72, (5), 610-617.

Howard, G.S., & Maxwell, S.E. Correlation between student satisfaction and grades: A case of mistaken causation? Journal of Educational Psychology, 1980, 72, (6), 810-820.

Howard, G.S., & Maxwell, S.E. Do grades contaminate student evaluations of instruction? Research in Higher Education, 1982, 16, (2), 175-188.

Marsh, H.W. The influence of student, course, and instructor characteristics in evaluations of university teaching. American Educational Research Journal, 1980, 17, (1), 219-237.

Marsh, H.W. Validity of students' evaluations of college teaching: A multitrait, multimethod analysis. Journal of Educational Psychology, 1982, 74, (2), 264-279.

Marsh, H.W., & Overall, J.U. Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. Journal of Educational Psychology, 1980, 72, (4), 468-475.

Overall, J.U., & Marsh, H.W. Relative influence of course level, course type, and instructor on students' evaluation of instruction. Paper presented at the meeting of the American Educational Research Association, Boston, April 1980. (ERIC Document Reproduction Service No. ED 187 176)